Adaptation of Thematic Apperception Test in Pakistan

Faiza Ali and Roomana Zeb
University of Peshawar

Thematic Apperception Test is a widely used projective measure of personality assessment. The present research adapted 20 of the most useful TAT cards according to Pakistani culture. The cards were modified by changing the dressing and facial features of the characters showed in the pictures. The cards were given to clinical psychologists for qualitative item analysis. After the final approval both the original and adapted TAT version were applied on 54 respondents (n=34 normal and n=20 psychological patients). Nine judges interpreted TAT protocols according to Bellack's scoring system on original as well as adapted version. Validity of the adapted version was calculated through convergence between original and adapted versions of TAT cards. Reliability was estimated through percentage of agreement among judges by inter-rater method. The validity ranged from 29.83 for card 7GF to 42.40 for card 9 BM. The average reliability ranged from 9.73 for card 18 BM to 22.79 for card 1. Although the reliability and validity indices of the newly adapted cards were generally low due to the projective nature of TAT, still the adapted version of TAT can be used in clinical settings for diagnostic purposes as well for the assessment of normal population .

Keywords: thematic apperception test, inter-rater reliability, convergent validity

An individual's personality is the core of a person with which s/he interacts with others. It's a combination of traits that affects the actions, thinking, desires, feelings and behaviour of a person. Interest in the study of personality originated with Hippocrates in 370 BCE. Psychology encompasses personality with the help of many perspectives including psychoanalysis (unconscious), behaviourism (overt actions), humanism (growth potential) and cognitive (thinking). Among them, psychoanalysis (Schultz & Schultz, 2016) focuses on the latent aspects of personality. The main assumptions of psychodynamic theory are: i) the basic determiner of our behaviour is the unconscious; ii) which originate from our painful childhood experiences; iii) and determines our behaviour; vi) and that personality is mainly developed during childhood.

Freud used free association to reach to the hidden aspects of a person's life, that led to the development of "projective hypothesis" that is our unconscious is revealed with the help of unstructured stimuli. Projective techniques are gaining support among clinicians due to improved standardization. Although there are many interpretations of projective techniques, most clinicians strongly believe in its effectiveness of revealing the hidden aspects of personality (Cordón, 2005). Murray was also impressed by Freud's notion of unconscious; his theory of personality has borrowed many concepts from psychoanalysis (Hall, Lindzey, Loehlin, & Manosevitz, 1985) that is why he decided to work on development of projective measure.

Correspondence concerning this article should be addressed to Dr. Roomana Zeb, Assistant Professor, Department of Psychology, University of Peshawar, Email: roomazeb@uop.edu.pk

Projective tests can be in the form of drawing, sentence completion, inkblots and semi structured sketches; Murray in 1930 (as cited in Anderson, 1999) worked on Thematic Apperception Test (TAT). This technique ask subjects to write a story on semi structured pictures, which in turn reveals their hidden motives, desires, conflicts and the like. In the making of TAT, Murray and Morgan spent 1930s in selecting pictures for TAT from different magazines (Morgan, 2002). After three times revision (series A, B and C), a final set of images (series D; which are now in use) that contain 31 cards was finalized. Initially Morgan was given the authorship credit of TAT, however, she refused to receive it on the final published instrument as she was not able to answer many letters having queries about TAT. TAT cards depict people in different actions or human interactions. Different TAT cards can be used for the assessment of male, female, children, adolescents and adults. Generally 10 TAT cards are administered in a single session to a subject which can elicit meaningful hidden aspects of personality. Individual's spontaneously written stories are scored and interpreted according to the content of these stories. Murstein (1963) has provided different content and quantitative scoring and interpretive analyses for thematic techniques. Other quantitative analyses also have been developed by researchers. Although been useful for research purposes these techniques bear little clinical application. That is why most clinicians prefer qualitative approaches for interpretation.

For administering TAT cards the person's ability to relate easily to the selected cards is also important. Murstein, (1968) suggest that the practitioner should decide most suitable 10-12 TAT cards, according to the subject. For example, age and gender visible in the TAT card is important to establish a connection with it. Generally cards 1, 3BM, 7GF, 8BM, 12M, 13B, 14 and 17BM are considered for children between 7 and 11 years of age. Productive TAT cards for adolescents are card number 1,2,5, 7GF, 12F, 12M, 15, 17BM, 18BM and 18GF. The practitioners can add one or two more cards which are considered essential for a better understanding of the subject. Like, for a depressed person, TAT pictures associated with depression and suicide can be added. Similarly, Symonds (1939) suggests ambiguous pictures with minimal detail as the most useful, specifically to overcome the effect of sad and dark tones noted in many TAT cards. Similarly, Goldfried and Zax (1965) found that ambiguous cards have usually been rated as less sad.

There are many scoring approaches of TAT, regardless of the scoring system; TAT helps reveal the individual's dominant drives, emotions, traits and conflicts. Murray's (1971), scoring is based on identifying significant interpersonal "*needs*" (determiners of behavior within the person) and "presses*"* (environmental determinants of behavior). In his scoring system the need is ranked on five point scale from 1 to 5 depending on the strength, duration, frequency and importance of needs as manifested in the story. Presses are also scored for intensity and frequency on a 5 point scale. As compared to Murray's classification, Bellack scoring focuses more on projection of defence mechanisms, ego functioning, interpersonal relations and the like.

Keeping in view the effectiveness of TAT, it's adapted versions are available for different populations. For example, for Negro subjects TAT was adapted on the assumption that blacks will identify easily with similar looking figures than of white subjects. Consequently, 23 of the TAT pictures were modified by substituting black with white figures (Thompson, 1949). Similarly, in Indian museum of Calcutta an adaptation of the TAT also took place. The pictures were adapted according to Indian culture (Chaudhary, 1960).

Rationale and Importance of the study

Cross cultural psychology suggests that the tests may not yield the same measure when used for assessment in a different culture. Every culture has specific traditions, customs, norms, believes etc which may affect responses on the measurement instruments. Matsumoto (as cited in Durrani, Mahmood & Saleem, 2017) stated that social and cultural aspects are an inseparable part of human development. Similarly, Maqsood, Hanif, Rehman, and Glen (2012) endorsed the need of assessing the measurement models developed in the West to see their generalizability across cultures.

Poortinga and Van der Flies (1988) suggested three things for cross-cultural application of scales, 1) the ability or characteristic assessed should have similar meaning in each culture; 2) it should organize behaviours in the same manner in each culture and, 3) test scores should have same meaning in each culture.

To achieve test equivalence in a different culture of Pakistan, TAT cards should be adapted by implementing appropriate changes. Usually people report difficulty in appropriately relating to TAT pictures due to dissimilar dressing, features and social setup. It is referred as "distancing" by clinical psychologists. Keeping in view this aspect, the present study aims to make TAT more useful and productive measure in general and specifically as an assessment tool in clinical settings.

As the present research focused more on helping clinicians reach accurate diagnosis, Bellak system of interpretation was used for scoring instead of Murray's. The present study aims to make TAT more relevant for Pakistani population, helping in selecting the most useful cards for assessment.

Objectives
1. Adaptation of TAT cards according to Pakistani culture
2. To establish psychometric properties of adapted TAT cards

Method

Sample
The present study consisted of 54 respondents, including diagnosed psychiatrist patients (n=20), and psychologically healthy (n=34). The sample varied in age between 16 to 50 years, with 50 % males (n=27) and 50 % females (n=27). As the sample was selected from government hospitals, most of the sample was less educated and belonged to lower to middle socioeconomic status. Convenience sampling technique was used to extract he sample from Peshawar, Abbotabad, and Dera Ismail Khan (cities comprising different division of Khyber Pakhtun Khawa).

Inclusion / exclusion criteria for the selection of clinical sample consisted of selecting individuals suffering from minor psychological disorders (like depression, anxiety). Whereas, severely disturbed individuals were excluded as they could not narrate stories on TAT cards. Individuals who already had responded to TAT were also excluded from the sample. The clinical sample was taken from different outpatient and inpatient departments of the hospitals situated in Peshawar, Abbotabad and Dera Ismail Khan

Instruments
Thematic Apperception Cards. . Original Thematic Apperception Test (Morgan, 1935) consists of 31 cards showing different situations including one blank card. Individual narrates

a story on the pictures shown in the card and the hidden motives, conflicts, defenses etc. are measured with the help of the stories. In the present study to establish the psychometrics of the adapted TAT, 20 of the original TAT cards were used i.e., card 1, 2, 3BM, 4, 5, 6BM, 6GF, 7BM, 7GF, 8BM, 8GF, 9 BM, 9 GF, 10, 12M, 12F, 13MF, 15, 18BM and 18GF. For scoring, Bellak system was utilized for interpretation.

Adapted Thematic Apperception Cards. In the present study the adapted 20 TAT cards were used (card number 1, 2, 3BM, 4, 5, 6BM, 6GF, 7BM, 7GF, 8BM, 8GF, 9 BM, 9 GF, 10, 12M, 12F, 13MF, 15, 18BM and 18GF) which were suggested as the most productive ones by the clinical psychologists for personality assessment. TAT stories were interpreted on Bellak scoring system.

Procedure

The process of adapting TAT cards initially started with consulting clinical psychologists who suggested frequently preferred cards for adaptation. Based on their suggestion; card 1, 2, 3BM, 4, 5, 6GF, 6BM, 4, 5, 6BM, 6GF, 7BM, 7GF, 8BM, 8GF, 10, 12M, 12F, 13MF, 15, 18BM and 18GF were selected for adaptation. After two unsuccessful trials the third artist remade the TAT cards effectively according to Pakistani culture. These cards were then presented to the panel of experts who assessed them for accurateness. They suggested minor changes which were incorporated in the sketches and the cards were shown to them again. They considered it appropriate for Pakistani population so the data collection was started. After getting informed consent one card at a time was shown to the subject and the subject narrated a story about it. With the time gap of one day, both versions of TAT cards were administered on the same subjects, the original and the adapted respectively. To break the ordering effect, original and adapted TAT cards were randomly administered. The cards left were shown to the person on the next day, but if the patient felt tired or not willing to respond further the process continued for three or more days to evaluate them on all cards. For the assessment of the TAT stories initially six clinicians were consulted to evaluate responses on both the versions of TAT. Each clinician had to interpret the protocols of 20 subjects. Assessment of projective measure is a difficult and time consuming task, due to which instead of six judges evaluating all 60 protocols, a set of two judges was made for the assessment of inter-rater reliability, i.e., 20 same protocols were assessed by judge one and two; 20 another by judge three and four and similarly the last 20 by judge five and six. The number of judges was increased afterwards due to problems faced in interpretation.

Details of how these cards were made and applied are as follows. Initially an artist was searched who can redraw the sketches in a culturally relevant manner. The researcher found one local artist from D.I.Khan and had a meeting with him to explain the concept behind this remaking of cards. He was explained to keep the original drawing a standard in which only features and dressing of the subjects need to be changed according to Pakistani population. After four months when the adapted version of cards was presented, it was the exact copy of the original without any cultural changes implemented in them. Although the first set of pictures did not fulfil the purpose of the research, but as the artist had good drawing skills, he was again explained the whole procedure. After another four months the second set presented consisted of the artist's own artistry in the cards, which made it very different from the original ones. As the artist was not getting what is standardization of a psychological test, another artist was searched. This time the department of Arts and Design of University of Peshawar was contacted where the chairman suggested the names of the two female artists. One of them refused, while the second one agreed. During another four months duration, researcher frequently asked the artist to send the soft copy of the cards, to make sure this time instructions are followed appropriately. It helped in correcting the minor mistakes timely. The complete set of 20 adapted cards was then presented to the judges

for evaluation. Three judges evaluated the drawings and few corrections suggested by them were incorporated in the sketches, after which the cards were considered ready for application.

Data collection phase had its own difficulties. People who initially agreed to take part in the study refused to participate in the middle of the process of data collection due to 40 cards to respond (20 original, 20 adapted). Due to an ethical guideline one cannot be forced to continue participation in the study, all that incomplete data became useless. Data collection from individuals with psychological problems was even tougher. Different hospitals and psychiatric clinics of Peshawar, Abbottabad and D.I.Khan were consulted for this purpose. The authorities allowed the researcher but individuals with psychological problems hardly agreed on making 20 stories on each version. As all the cards could not have been applied on a single day, it required taking one day gap between the administrations of both versions. Many of the patients did not cooperate on the second day.

Interpretation of the TAT cards was the most hideous task to carry. Initially 6 judges were supposed to serve the purpose, but when each judge received 20 protocols, four of them refused, although they had committed earlier. One judge interpreted the complete 20 protocols, another judge interpreted only 6 forms out of 20 protocols. Due to this problem more clinical experts were searched for interpretation of TAT cards. New judges sought also did not agree because of the time frame left for the study (this paper is extracted form an MPhil work and less time was left for the submission of thesis) and plenty of forms. Finally the data was now distributed among seven judges who agreed to interpret TAT cards. Due to the unwillingness of judges for interpreting 20 protocols, it was reduced to 10 protocols i.e., each judge was given a set of ten protocols. Judges were given 15 days for interpretation due to leaving some time for the compilation of results, but the judges interpreted the set of protocols according to their convenience. Each judge took two and half months to interpret the protocols. At the end every judge returned 2 or 3 forms without interpretation. They excused that they cannot manage to interpret the whole set of ten protocols because of their personal responsibilities and time constraint.

Results
Estimation of Validity. Convergent validity of TAT was estimated through calculating similar indicators on both versions of TAT and then converting it to percentage of agreement among judges. Two separate judges independently filled the Bellack's form on both original and adapted cards. Total indicators found on both the versions, helped in constituting similarity index between the original and adapted version, i.e., counting similar indicators found by a single judge on both the versions. The estimated validity by a single judge for a particular card was found by dividing similar indicators on total indicators found. The same procedure was adopted for the second judge. For each of the card then an average validity was estimated by summing the indicators of the two judges and dividing it by 2.

Table 1

Average validity of different TAT cards computed through convergent validity between original and adapted TAT cards by the judges

Cards	Validity index
1	31.29
2	27.86
3BM	28.60
4	27.78
5	28.68
6BM	25.93
6GF	23.95
7BM	28.56
7GF	29.83
8BM	30.64
8GF	26.39
9BM	42.40
9GF	29.59
10	33.08
12M	29.15
12F	29.43
13MF	26.81
15	26.27
18BM	27.75
18GF	28.28

Table1 shows the average validity of each TAT card with lowest validity 23.95 for card 5 and highest 42.40 for card 9BM.

Estimation of reliability. Reliability indices were calculated through inter-rater method, i.e., two separate judges interpreted the same adapted card. Consensus between the judges on indicators found for a given card gave higher reliability.

Table 2

Average inter-rater reliability of different TAT cards computed through percentage of agreement between judges

Cards	% of agreement
1	22.79
2	21.42
3BM	16.97
4	17.63
5	12.25
6GF	14.12
6BM	18.74
7GF	17.96
7BM	15.54
8GF	12.15
8BM	12.61
9GF	14.12
9BM	18.55
10	18.11
12F	15.49
12M	14.06
13MF	19.05

15	12.55
18GF	11.68
18BM	9.73

Table 2 shows the inter-rater reliability of adapted cards by calculating all the similar indicators filled by two judges on adapted version. The highest reliability index is 22.79 (card 1) and lowest 9.73 (card 18BM).

Discussion

As Thematic Apperception Test is a widely used projective measure of personality. One problem it poses when applied in other cultures is "distancing", that is due to differences in features and dresses of TAT pictures, individuals find it difficult to relate with them appropriately (Habib, 2016, Personal communication). Keeping in view this limitation it was decided in the present study to adapt TAT for Pakistani population. As has mentioned in the procedure section, out of the 30 TAT cards, 20 cards that were suggested by clinicians were selected for adaptation.

Table 1 shows that both cards 6GF and 6BM had the lowest validity index 23.95 and 25.93 respectively. As these pictures are less structured the respondents did not narrate detailed meaningful stories on both original and adapted version of these cards. Looking for the highest validity indices, card 9BM scored the highest (42.40) followed by card 1 (31.29). It suggests that these two cards are more meaningful for the respondents in our culture and should be incorporated in the set of cards to be administered on the respondents.

Table 1 shows the validity indices which are not high for most of the cards. In a projective test, we can come across such problems as in the present study quantification was applied on the qualitative data. Low validity and reliability is the common problem of projective measures (Riaz, 2008) as it is not necessary that on both the original and adapted versions similar stories, with similar, needs or defence mechanisms will be achieved. It is likely that a similar but not identical defence is seen in the adapted than original card, which lowers down the validity. As in the present research long form of Bellack's system was used for interpretation, finding similarity between numerous indicators was difficult to achieve.

Another reason for the low validity is that the respondents had to answer 40 cards in total, knowing that they narrated shorter stories, which the judges did not find useful, hence left may indicators blank which resulted in loss of the data. Lack of interest on the part of judges was another reason, as they did not assess all cards with same vigour which resulted in lower validity indices.

Table 2 shows reliability index of the 20 adapted TAT cards. As like validity, card number 1 also showed higher reliability. Similarly, card 2 also had higher reliability than the other cards. Inspecting the average reliability of Card 1 (22.79) showed subject no 24's responses yielded a zero consensus among the two raters, whereas, the highest agreement was found for subject 8's responses i.e., 38.90. Average reliability was 21.42 for card 2 in which for subject 3 the judges showed lower consensus, whereas, for subject 7 there was highest consensus among the judges, i.e., 36.84.

Lilienfeld, Wood, and Garb (2000) also stated in their meta-analysis of projective measures that generally these measures are low on psychometric properties. As in the present study results indicated lower reliability estimates than the validity due to the dependence on finding consensus in the responses of two separate judges. Various reasons are responsible for variation among the interpretations of judges in projective measures, e.g., differences in the

experience and training. Similarly, in the present research when the researcher was calculating reliability, similar but non-identical defences were scored different from each other which resulted in lower reliability of all the cards. It is evident to face such problems in calculation of reliability of projective measures as the data is narrative that is converted in quantification. Similarly, due to the existence of different scoring systems with their own limitations it is difficult to reach to higher reliability indices. In practice, mostly experts trust their instincts in interpretation to make their judgments instead of strongly adhering to a scoring system. As different scoring systems yield different instructions, requiring diversity of numbers, categories in arrangement and presentation of cards from one individual to the other (Geiser & Stein, 1999; Karon, 2000), which generally results in low reliabilities. Similarly the problems can also occur due to the difference between administrations. When under the same administration the scoring of stories can be different, variations are more likely to occur under different administrations, hence reliability is expected to be low (Entwisle, 1972).

Lundy (1985) estimated test retest reliability of .56 and .48 respectively for the need for association and accomplishment when assessed subjects on story telling with a time gap. As the consistency of test-retest was low, it is stated that these consistencies only revealed the stability of memory instead of the stability of personality traits on TAT.

Limitations
- As the judges needed more time for interpretation, they skipped some cards which in turn lowered the validity and reliability indices.
- Similarly, some of the judges interpreted only gender specific cards when interpreting male and female respondents.
- Some judges interpreted only adapted cards while leaving the original TAT cards. Similarly, the length of interpretations differed among the judges.
- Lack of finances was a hindrance in interpretation phase.
- Due to responding to 40 cards, subjects wrote shorter stories which made the interpretation difficult for the judge.

Suggestions
- Keeping in view the above limitations the following suggestions are put forward:
- Finances must be provided to motivate the judges for assessing TAT responses.
- More productive responses could be gathered by applying only gender specific cards on the respondents instead of all the adapted ones.
- Murray scoring is easy in quantification, so future researchers can use it for establishing the reliability of TAT.
- More than two judges should be used while calculating inter- rater reliability.

References

Anderson, J. W. (1999). Henry A. Murray and the creation of the Thematic Apperception Test. In L. Gieser & M. I.   Stein (Eds.), *Evocative images: The Thematic Apperception Test and the   art   of   projection* (pp.   23–38).   American   Psychological Association. https://doi.org/10.1037/10334-002

Bellak, L. (1975). *The TAT, CAT, and SAT in clinical use*. (3rd ed.). New York: Grune & Stratton.

Chaudhary, U. (1960). *Indian adaptation of TAT*. New Delhi: Manasayan.

Cordón, L. A. (2005). *Popular psychology: An encyclopedia*. Westport, Conn: Greenwood Press. pp. 201–204.

Durrani, S. M., Mahmood, Z., & Saleem, S. (2017). The development and validation of Temperament Scale for university students. *FWU Journal of Social Sciences,11*(1),264-275.

Entwisle, D. R. (1972). To dispel fantasies about fantasy-based measures of achievement motivation. *Psychological Bulletin, 77*, 377-391.

Exner, J. E. (1976). Projective techniques. In I. B. Weiner (Ed.), *Clinical Methods in Psychology.* John Wiley & Sons.

Geiser, L., & Stein, M. I. (1999). Evocative images: *The Thematic Apperception Test and the art of projection*. Wasington, DC: American Psychological Association.

Goldfried, M. R., & Zax, M. (1965). The stimulus value of the TAT. *Journal of Projective Techniques and Personality Assessment, 29*, 46-67.

Hall, C. S., Lindzey, G., Loehlin,J. C., & Manosevitz, M. (1985). *Theories of personality*. Ohio: Wiley

Karon, B. P. (2000). The clinical interpretation of Thematic Apperception Test: A reexamination of statistical versus clinical prediction. *Professional Psychology: Research and Practice, 31*, 230-233.

Lilienfeld, S. O., Wood, J. M. , & Garb, H. N. (2000). The scientific status of projective techniques. *Psychol Sci Public Interest*. *1*(2), 27-66. doi: 10.1111/1529-1006.002.

Lundy, A. (1985). The reliability of Thematic Apperception Test. *Journal of Personality Assessment, 49*, 141-145.

Maqsood, A., Hanif, R., Rehman, G., & Glen, W. (2012). Validation of the three-component model of organizational commitment questionnaire. *FWU Journal of Social Sciences, 6* (2), 135-145.

Morgan, W. (2002). Origin and history of the Thematic Apperception Test. *Journal of Personality Assessment, 79*(3), 422-445.

Murray H. A. (1938) *Explorations in personality*. New York: Oxford University Press.

Murray, H. A. (1971). *Thematic apperception test manual*. Cambridge, MA: Harvard University Press.

Murstein, T. (1963). *Psychological testing*. Palstation, New York: Lumare Associates.

Murstein, B. I. (1968). Effects of stimulus, background, personality and scoring systems on the manifestations of hostility on the TAT. *Journal of Counseling, and Clinical Psychology, 32,* 335-365.

Poortinga & Van der Flies (1988). The meaning of item bias in ability tests. In S. H. Irvine & J. W. Berry (eds.), *Human abilities in cultural context*. Cambridge: Cambridge University Press, pp. 166–183

Riaz, M. N. (2008). *Test construction: Development and standardization of psychological tests in Pakistan*. Islamabad: Higher Education Commission.

Schultz, D. P., & Schultz, S. E. (2016). *Theories of personality*. Cengage Learning.

Symonds, P. M. (1939). Criteria for the selection of pictures for the investigation of adolescent fantasies. *Journal of Abnormal and Social Psychology, 34*, 271-274.

Thompson, C. E. (1949). *Thematic Apperception Test: Thompson modifications*. Cambridge, M. A: Harvard University Press.

Appendix

Adapted TAT Cards